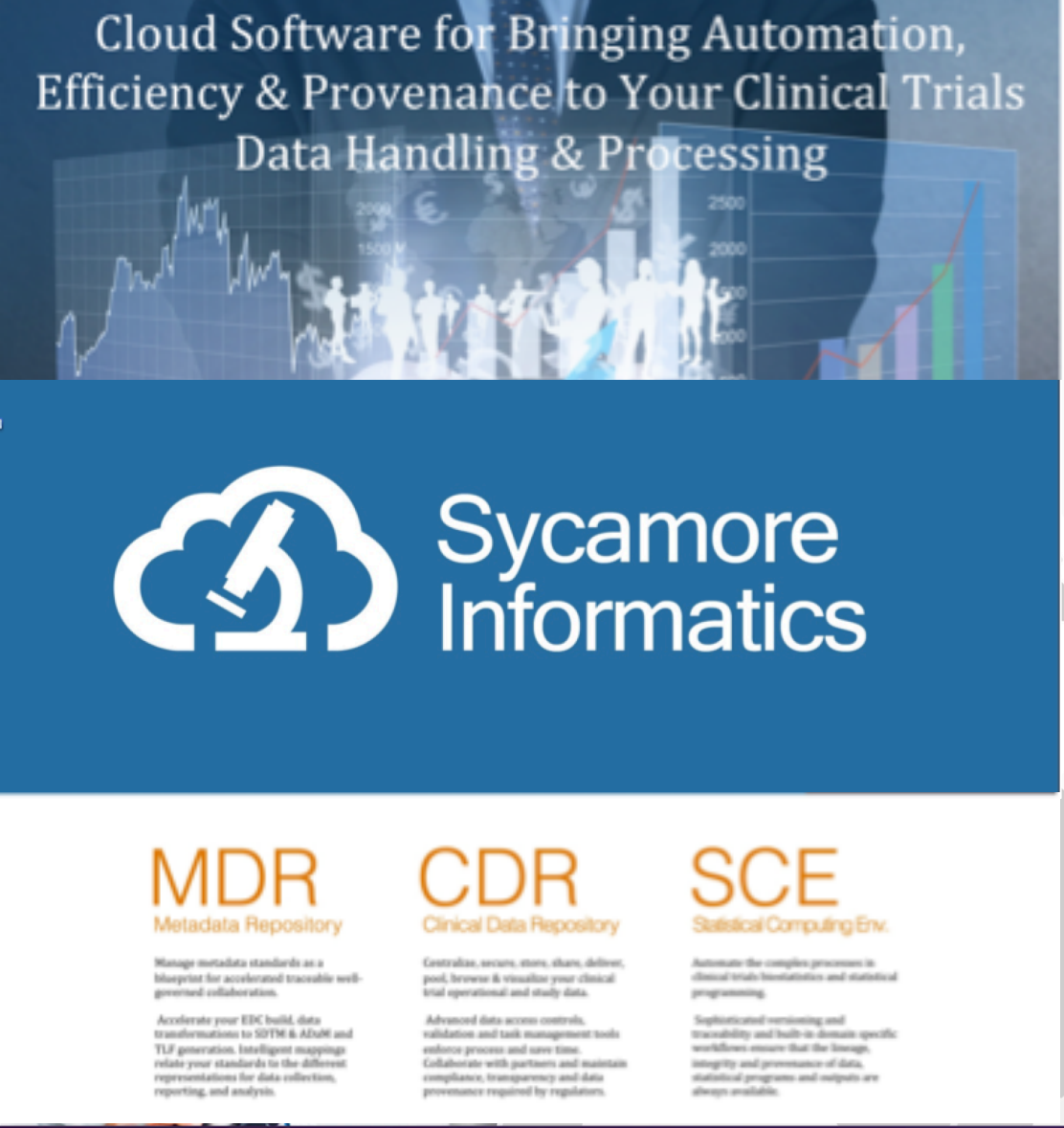


To Version or Not (When) to Version

Joel Hoffman, PhD

Topics

- Frame of Reference and the business problems
- Published comments on version and version control
- Use Cases
  - Managing Data Collaboration
  - Managing Analysis and Reporting Deliverables
  - Managing Regulatory Interactions
- Q&A and Summary

Frame of Reference and the Business Problem

- For analysis and reporting of clinical trial data:
  - Traceability – i.e., transparency into the analysis and reporting process
  - Identify "the output" used for decision making and regulatory submissions
  - Replicate the generation of "the output" to respond to regulatory, scientific, and business queries
  - Access storage and computation resources in place
  - Available where, when, and to whom it is needed
  - Efficient to use
  - Minimize change to business processes (SOPs and WPs)

IS IT TRUE WHAT THEY SAY ABOUT VERSIONS – THE WHAT & HOW

Hopkins, Duke and Dubman

- Version control of the output includes all the input data, production libraries, and program logs.
  - Dependency management of inputs and outputs is necessary because of the complexity of multilevel steps and the often ad hoc nature of data analysis.

Hopkins, Duke and Dubman

- Ease of use is critical to user acceptance of the new working environment.
- Early program development and the iterative changes associated with making a working program stable could be done outside the SCE to avoid version control when it may not be most useful. Once programs start producing output shared with a project team, business rules typically kick in and require that the programs reside in the SCE. Then further changes are subject to version control.

Hopkins, Duke and Dubman

- Must be able to run multiple statistical packages, e.g., SAS, R, S-PLUS, etc and multiple versions of supported packages due to the fact that the clinical program life cycle will often run longer than a particular version of a software product

Traceability and Data Flow – PhUSE October, 2014

- two types of traceability each requiring their own versions
  - Metadata traceability as “describing (via metadata) the algorithm used or steps taken to derive or populate an analysis value from its immediate predecessor”
  - Data-point traceability as enabling “the user to go directly to the specific predecessor record(s)”

Wayne Woo: BUILDING A CONTROLLED STATISTICAL PROGRAMMING ENVIRONMENT

- Protocols, analysis plans, programming specifications and mock tables represent requirements generated from the planning stage and need to be safely stored and versioned
- When validation is completed, the program should be elevated to a production status and versioned.
- When a file is made "production", it must be versioned so that in the future a particular output can be traced to a particular set of source code.

Wayne Woo: BUILDING A CONTROLLED STATISTICAL PROGRAMMING ENVIRONMENT

- As stable versions are produced, it is desired to permanently archive these versions.
- There should be monitoring by managers and a mechanism to periodically evaluate conformance to the folder structure standard. To allow for program validation and version control, separate folders for development, testing, and production are needed. There must be a process and mechanism to move files upward from development to testing and ultimately production. This "file promotion" should be recorded.

Wayne Woo: BUILDING A CONTROLLED STATISTICAL PROGRAMMING ENVIRONMENT

- If changes are needed later, programmers must be able to retrieve the current production version and use it as the basis to make revisions
- The software should be installed via a formal installation qualification script to document the process and for reproducibility if hardware changes. There should be maintenance of versions, …

SUMMARY OF COMMENTS AND ADDITIONAL EXPERIENCES

Summary of Functions

- A linked set
- Ease of Use
- Should “kick-in”
- Promotion to production
- System should be installed with formal IQ
- ????

Summary of Functions

- A linked set
- Ease of Use – Shouldn't get in the way
- Should "kick-in"
- Promotion to production
- System should be installed with formal IQ
- Access to resources – Accessible by Resources

Summary of What to Version

- Includes:
  - Programs, data, output, libraries
  - Protocol, SAP, Mock Table (Requirements for SDLC)
  - System (formal IQ)
  - Software (OS, SAS, R, …)
  - Permanent Archive
  - Metadata and Data Points
  - Dependencies
  - ???

Summary of What to Version

- Includes:
  - Programs, data, output, libraries
  - Protocol, SAP, Mock Table (Requirements for SDLC)
  - System (formal IQ)
  - Software (OS, SAS, R, …)
  - Permanent Archive
  - Metadata and Data Points
  - Dependencies
  - Specifications

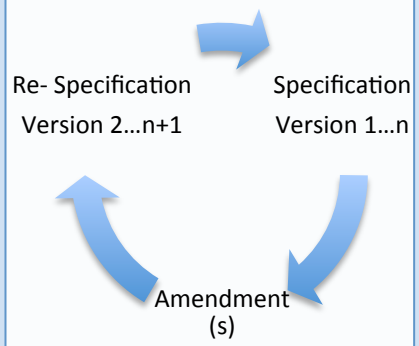
USE CASE 1: MANAGING DATA COLLABORATION

Managing Data Collaboration

Industry, Company, CRO Standards

Data Collection Instruments

- eCRF/DCI



- Supplied (as per Spec)

Reporting

- Supplied SDTM

Study	•As per Spec •Safety Review
Cross Study	•Structure •Value List
Submission	•OpenCDISC •DEFINE.XML •SDRG

- Final SDTM (as per Spec)

Analysis

- Supplied ADaM

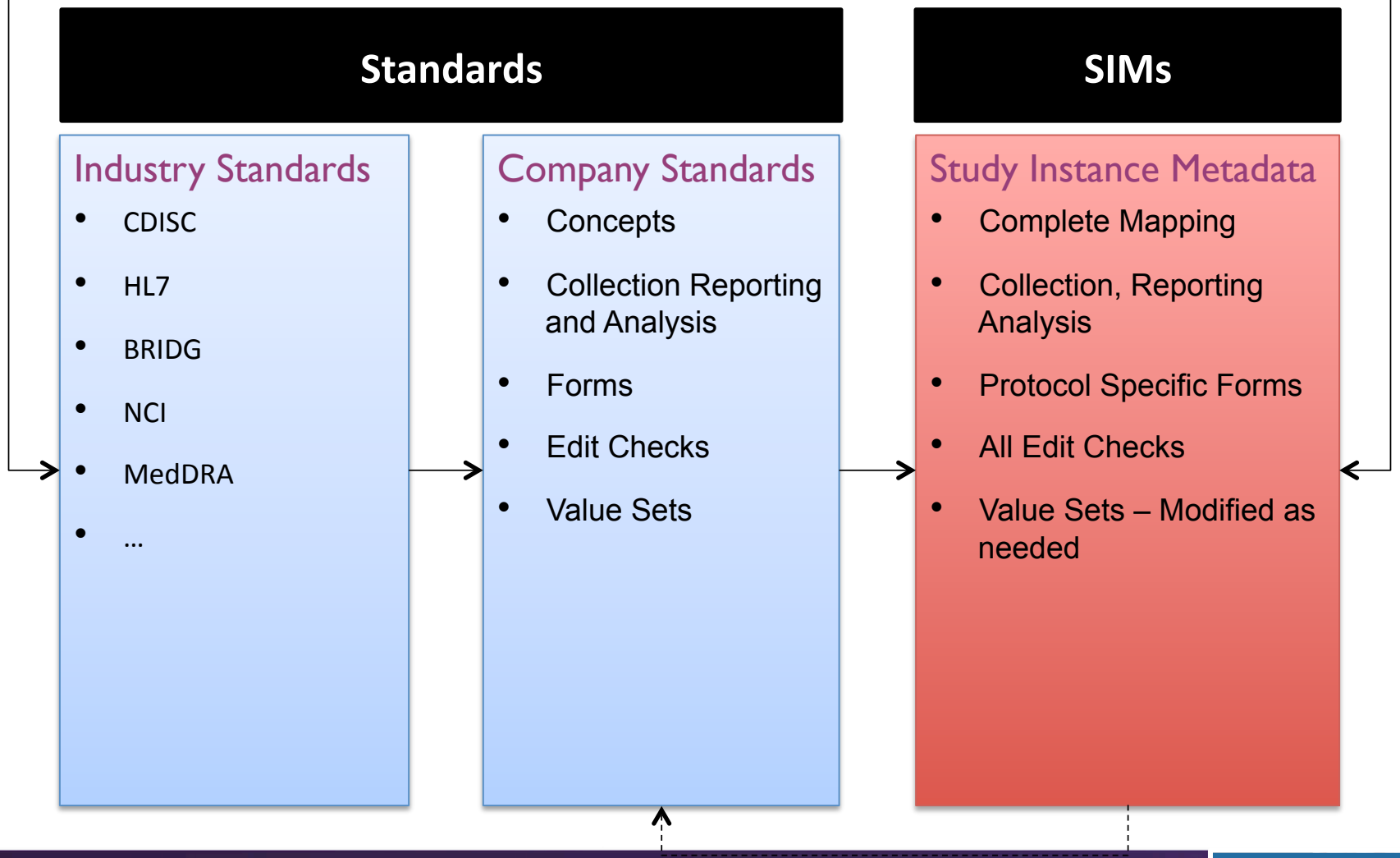
Study	•Interim and Final Report
Cross Study	•Structure •Value List •Algorithms
Submission	•DEFINE.XML •ADRG

- Final ADaM (as per Spec)

"Standards" and "SIMs"

Files from Source

Protocol, SAP, DQP



Managing Data Collaboration

- Data Collection Instrument
  - Standard eCRF Spec – Study eCRF Spec
  - Standard eCRF – Study eCRF
  - Standard Value List – Study Value List
- Reporting
  - Standard STDM – Study SDTM
  - Source → SDTM Map/Specification
  - Source → SDTM Code
- Analysis
  - Standard ADaM – Study ADaM
  - SDTM → ADaM Map/Spec
  - SDTM → ADaM Code

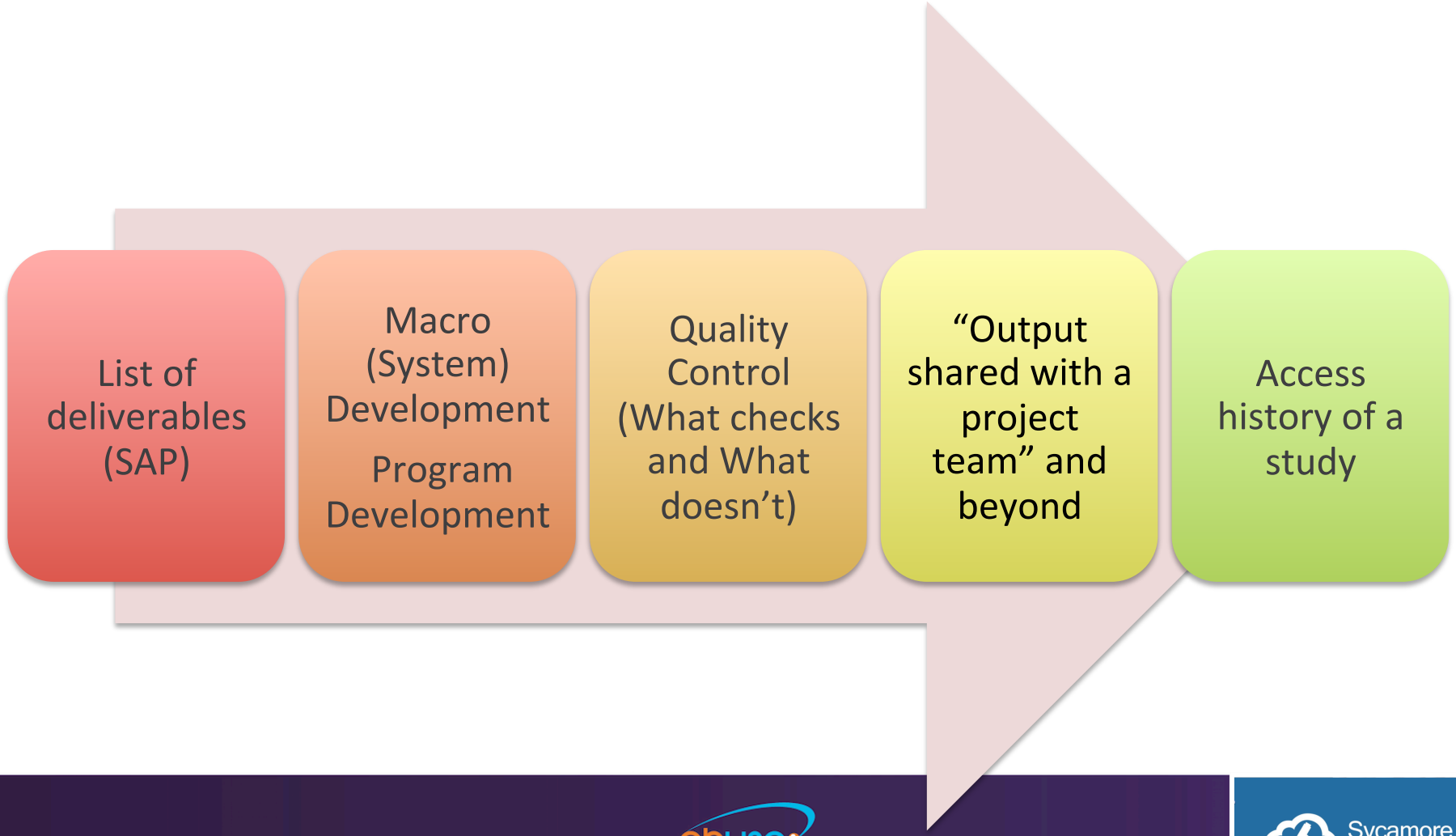
Metadata Repository (MDR)

Statistical Computing Environment (SCE)

Clinical Data Repository (CDR)

USE CASE 2: ANALYSIS AND REPORTING DELIVERABLES

Managing Analysis and Reporting Deliverables



USE CASE 3: INTERACTIONS WITH REGULATORY AGENCIES

Interactions with Regulatory Agencies

"If changes are needed later, programmers must be able to retrieve the current production version and use it as the basis to make revisions"

	Number of Drugs	Sales 2014	Average per drug per day
Drugs Current on Patent*	12233	$259,092,876,285**	$58,026.94
Highest Sales	20	$85,000,000,000***	$11,643,835.62
Example	1	$3,000,000,000	$8,219,178.08

*FDA Orange Book Sept 2015
** Kaiser Family Foundation
*** The Statistical Portal

Q&A and Summary

Joël Hoffman

jhoffman@sycamore informatics.com

References

- Hopkins, Duke, Dubman, "Statistical Computing Environments and the Practice of Statistics in the Biopharmaceutical Industry", Drug Information Journal, Vol. 44, 2010
- BUILDING A CONTROLLED STATISTICAL PROGRAMMING ENVIRONMENT, Wayne Woo, Novartis Vaccines & Diagnostics, Cambridge, MA, WUSS 13 Paper
- Traceability and Data Flow – PhUSE October, 2014


phuse
The premier community for people working in the biometric area
@PhUSETwitta
PhUSE
www.phusewiki.org